* For regression models, we assume:

1) $\varepsilon_i \sim N(0, \sigma^2)$ $\qquad Y = f(x_i) + \varepsilon_i$
2) $Y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2)$ $\qquad$ linear / non-linear

* Generalised Linear Models ($Y_i \nsim N$):

* Exponential family of Distribution:

A random variable X (binom, Pois, N) that belongs to exponential family of distribution with a single parameter $\theta$ has the following PDF:

$$f(x, \theta) = \exp\{a(x) b(\theta) + d(x) + c(\theta)\}$$

$b(\theta)$ is called natural parameter
$a(x) = x$, the distribution is in canonical form

$$f(x, p) = \binom{n}{x} p^x (1-p)^{n-x} \quad ; \quad x = 0, 1, \ldots, n$$

$$= \binom{n}{x} \left(\frac{p}{1-p}\right)^x (1-p)^n$$

$$= \exp\left\{x \log\left(\frac{p}{1-p}\right) + n \log(1-p) + \log\binom{n}{x}\right\}$$

$a(x) = x$ $\qquad b(p) = \log\left(\frac{p}{1-p}\right)$

$c(p) = n \log(1-p)$ $\qquad d(x) = \log\binom{n}{x}$

$$f(x,\mu) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left( \frac{-(x-\mu)^2}{2\sigma^2} \right) \quad , \quad x \in \mathbb{R}$$

$$\mu \in \mathbb{R}$$

$$\sigma > 0$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left( \frac{-x^2 + 2x\mu - \mu^2}{2\sigma^2} \right)$$

$$= \exp\left\{ \frac{x\mu}{\sigma^2} + \left( \frac{-\mu^2}{2\sigma^2} - \frac{\log(2\pi\sigma^2)}{2} \right) - \frac{x^2}{2\sigma^2} \right\}$$

$a(x) = x$ $\qquad$ $b(\mu) = \frac{\mu}{\sigma^2}$

$d(x) = \frac{-x^2}{2\sigma^2}$ $\qquad$ $c(\mu) = \frac{-\mu^2}{2\sigma^2} - \frac{\log(2\pi\sigma^2)}{2}$

$$f(x,\lambda) = \frac{e^{-\lambda}\lambda^x}{x!} \quad , \quad x \in \mathbb{N}$$

$$= \exp\{ x \log(\lambda) - \lambda - \log(x!) \}$$

$a(x) = x$ $\qquad$ $b(\lambda) = \log(\lambda)$

$d(x) = -\log(x!)$ $\qquad$ $c(\lambda) = -\lambda$

$$f(x,\lambda) = \lambda e^{-\lambda x}$$

$$= \exp\{ -\lambda x + \log \lambda \} \quad , \quad x \geq 0$$

$a(x) = x$ $\qquad$ $b(\lambda) = -\lambda$

$d(x) = 0$ $\qquad$ $c(\lambda) = \log(\lambda)$

$$f(x,\alpha) = \frac{\beta^{\alpha}}{\ } x^{\alpha-1} e^{-\beta x} \quad , \quad x \geq 0$$

$\Gamma(\alpha)$

$$= \exp\{\alpha \log(\beta) - \log(\Gamma(\alpha)) + (\alpha-1)\log x - \beta x\}$$
$$= \exp\{(\alpha-1)\log x + (\alpha\log(\beta) - \log\Gamma(\alpha)) - \beta x\}$$

$a(x) = \log(x)$ $\quad$ $b(\alpha) = \alpha - 1$

$d(x) = -\beta x$ $\quad$ $c(\alpha) = \log\left(\frac{\beta^\alpha}{\Gamma(\alpha)}\right)$

$$f(x,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, \quad x \geq 0$$
$$= \exp\{\alpha \log(\beta) - \log(\Gamma(\alpha)) + (\alpha-1)\log x - \beta x\}$$
$$= \exp\{-\beta x + \alpha\log(\beta) + ((\alpha-1)\log x - \log\Gamma(\alpha))\}$$

$a(x) = x$ $\quad$ $b(\beta) = -\beta$

$d(x) = (\alpha-1)\log x - \log\Gamma(\alpha)$ $\quad$ $c(\beta) = \alpha\log(\beta)$

$$f(x,p) = \binom{n+x-1}{x} p^n (1-p)^x$$
$$= \exp\left\{\log\binom{n+x-1}{x} + n\log(p) + x\log(1-p)\right\}$$
$$= \exp\left\{x\log(1-p) + n\log(p) + \log\binom{n+x-1}{x}\right\}$$

$a(x) = x$ $\quad$ $b(x) = \log(1-p)$

$d(x) = \log\binom{n+x-1}{x}$ $\quad$ $c(x) = n\log(p)$

$Y \sim$ binom $\Rightarrow$ Logistic Regression

$Y \sim$ Poisson $\Rightarrow$ Poisson Regression

Credit Card Balance

| Income $X_1$ | Age $X_2$ | Credit Card Balance | Loan $Y$ |
| --- | --- | --- | --- |
| | | | Yes |
| | | | No |
| | | | Yes |
| | | | No |
| | | | ⋮ |

$Y \sim$ binom so we can fit a logistic regression

* How to fit a GLM?

Features : $X_1, \dots, X_p$ Sample size : $1, \dots, n$

Response : $Y$

$$f(Y_1, \dots, Y_n) = \exp\left\{\sum_{i=1}^{n} Y_i b(\theta) + \sum_{i=1}^{n} c(\theta_i) + \sum_{i=1}^{n} d(Y_i)\right\}$$

Example : $Y \sim$ binom

Suppose we have $(Y_i, X_i)$ where $Y_i$ is of the form $\frac{r_i}{n}$ where $r_i$ is # of loans given (success) & $n$ total # of people (trials) and prob $= p$

$$f(Y_1, \dots, Y_n) = \prod_{i=1}^{n} \binom{n}{Y_i} p_i^{Y_i} (1-p_i)^{n-Y_i}$$

$$= \prod_{i=1}^{n} \binom{n}{Y_i} \left(\frac{p_i}{1-p_i}\right)^{Y_i} (1-p_i)^{n}$$

$$= \exp\left\{\sum_{i=1}^{n} Y_i \ln\left(\frac{p_i}{1-p_i}\right) + \sum_{i=1}^{n} n \ln(1-p_i) + \sum_{i=1}^{n} \ln\binom{n}{Y_i}\right\}$$

$i=1$

Natural parameter: $\ln\left(\frac{p_i}{1-p_i}\right)$

$Y \sim Pois(\lambda)$

$$f(Y_1,\dots,Y_n) = \prod_{i=1}^{n} \frac{e^{-\lambda_i}\lambda_i^{Y_i}}{Y_i!}$$

$$= \exp\left(\sum_{i=1}^{n} Y_i \log(\lambda_i) - \sum_{i=1}^{n} \lambda_i - \sum_{i=1}^{n} \log(Y_i!)\right)$$

Natural parameter: $\log(\lambda_i)$

$Y \sim N(\mu, \sigma^2)$

$$f(Y_1,\dots,Y_n) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(Y_i-\mu_i)^2}{2\sigma^2}\right)$$

joint density of indep. $Y_i$'s

$$= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-Y_i^2}{2\sigma^2} + \frac{Y_i\mu_i}{\sigma^2} + \frac{\mu^2}{2\sigma^2}\right)$$

$$= \exp\left(\frac{-n}{2}\log(2\pi\sigma^2) + \sum_{i=1}^{n}\frac{Y_i\mu_i}{\sigma^2} + \sum_{i=1}^{n}\frac{\mu_i^2}{2\sigma^2} - \sum_{i=1}^{n}\frac{Y_i^2}{2\sigma^2}\right)$$

$$= \exp\left(\sum_{i=1}^{n}\frac{Y_i\mu_i}{\sigma^2} + \left(\sum_{i=1}^{n}\frac{\mu_i^2}{2\sigma^2} - \frac{n}{2}\log(2\pi\sigma^2)\right) - \sum_{i=1}^{n}\frac{Y_i^2}{2\sigma^2}\right)$$

Natural Parameter: $\frac{\mu_i}{\sigma^2}$

$Y \sim Exp(\lambda)$

$$f(Y_1,\dots,Y_n) = \prod_{i=1}^{n} \lambda_i e^{-\lambda_i Y_i}$$

$$= \exp\left(\sum_{i=1}^{n}\log(\lambda_i) - \sum_{i=1}^{n}\lambda_i Y_i\right)$$

Natural parameter : $\lambda_i$

* **Def** : Link Function

A link function is often regarded as the natural parameter.

We fit the model $\ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}$

we want to estimate $\mathbb{E}(Y_i) = p_i$

$$p_i = \frac{e^{\beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}}}{1 + e^{\beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}}}$$ , $p_i$ is the success prob.

$p_i > \frac{1}{2}$ give a loan $\qquad$ $p_i \leq \frac{1}{2}$ don't give a loan

This is called logistic model categorical / classification

$$\begin{cases} \ln\left(\frac{p_i}{1-p_i}\right) = x \\ \Rightarrow p = \frac{e^x}{1+e^x} \end{cases}$$

$$\log(\lambda_i) = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}$$

$\lambda_i = e^{\beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}}$ count data

$\hat{\beta} \neq (X^T X)^{-1} X^T Y$, cannot be computed like this!

$$\frac{\mu_i}{\sigma^2} = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}$$

$\mu_i = \sigma^2(\beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip})$

$$\mu_i = e^{(\beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip})}$$

$$\lambda_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}$$

## * Transformation : Linearizing a Model

- $Y = \beta_0 e^{x \beta_1} \Rightarrow \log(Y) = \log(\beta_0) + x\beta_1$

  $(Y, x) \Rightarrow (\log Y, \log x)$

- $Y = \dfrac{x}{\beta_0 x + \beta_1} \Leftrightarrow \dfrac{x}{Y} = \beta_0 x + \beta_1$

  $\Leftrightarrow \dfrac{1}{Y} = \beta_0 + \beta_1 \dfrac{1}{x}$

  $(X, Y) \Rightarrow \left(\dfrac{1}{X}, \dfrac{1}{Y}\right)$

- $\log(Y) = \log \beta_0 + \beta_1 \log x$

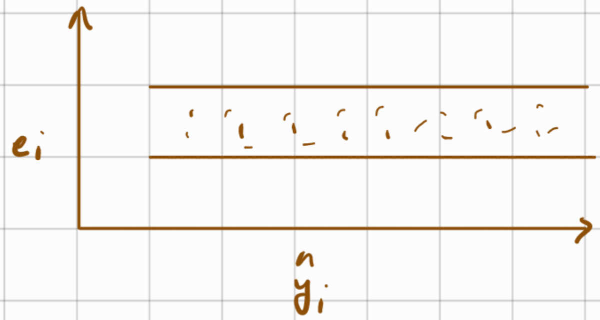
  $(X, Y) \Rightarrow (\log X, \log Y)$

- $Y = \beta_0 + \beta_1 \log x$

  $(X, Y) \Rightarrow (\log X, Y)$

- Box Cox Method : $\begin{cases} \dfrac{Y_i^{\lambda} - 1}{\lambda}, & \lambda \neq 0 \\ \log(Y_i), & \lambda = 0 \end{cases}$

$Y_i = f(x_i) + \varepsilon_i \qquad \varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$




$e_i$ : residual
$\hat{y}_i$ : fitted values

## Variance Stabilizing Transformation :

This is useful when the constant variance assumption is violated. Many times this is due to the fact that the response variance does not follow normal.

1) $Y \sim Pois(\lambda) \qquad E(Y) = Var(Y) = \lambda$

$$Y' = \sqrt{Y}$$

$$E(\sqrt{Y}) = e^{-\lambda} \sum_{y \geq 0} \sqrt{y} \, \frac{\lambda^y}{y!}$$

2) $0 < Y < 1 \qquad Y' = \sin^{-1}(\sqrt{Y})$

Taylor Series expansion :

$$f(x) = f(a) + f'(a)(x-a) + \frac{f''(a)(x-a)^2}{2!} + \cdots$$

where $f \in C^\infty([a-\varepsilon, a+\varepsilon])$

$Y \sim N(\mu, \sigma^2) \qquad g(\mu) = \sigma^2$

$Y' = f(Y) = f(\mu) + f'(\mu)(Y - \mu)$

$U = f(Y) = f(\mu) + f'(\mu)(Y - \mu)$ Taylor Order 1

$$
\begin{aligned}
Var(U) &= Var(f(Y)) \\
&= Var(f(\mu) + f'(\mu)(Y - \mu)) \\
&= (f'(\mu))^2 \, Var(Y) \\
&= (f'(\mu))^2 \, g(\mu) \\
&= 1
\end{aligned}
$$

$\Rightarrow f'(\mu)^2 = \dfrac{1}{g(\mu)}$

$\Rightarrow f'(\mu) = \dfrac{1}{\sqrt{g(\mu)}}$ , so with this choice, the variance is not constant

* Weighted Least Square method :

Used for "non constant variance" problem

$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$

$y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2)$

$loss_{OLS} = \sum e_i^2 = \sum (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$

$Wloss_{WOLS} = \sum w_i e_i^2 = \sum (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 w_i$ , $w_i \propto \dfrac{1}{\sigma^2}$

Normal Equation :

$\dfrac{\partial S}{\partial \beta_0} = 0 \Rightarrow \sum w_i Y_i = \hat{\beta}_0 \sum w_i + \hat{\beta}_1 \sum w_i X_i$

$\dfrac{\partial S}{\partial \beta_1} = 0 \Rightarrow \sum w_i X_i Y_i = \hat{\beta}_0 \sum w_i X_i + \hat{\beta}_1 \sum w_i X_i^2$

## * Logistic Regression :

$Y \sim binom(n, p)$ $\quad\quad Y = \{0, 1\}$

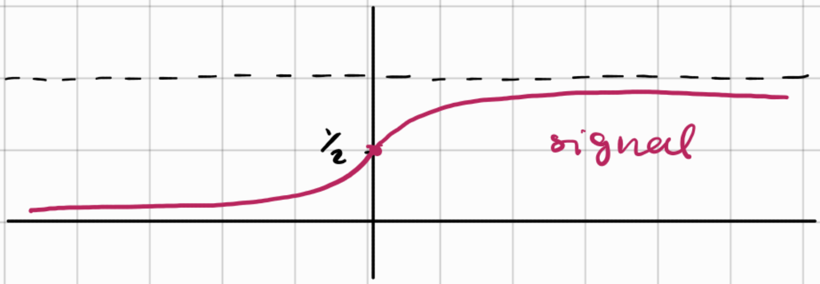
Based on pictures of cats and dogs we classify them 0 (cat) 1 (dog), from the pictures we can extract information like size of nose and ears. We can use logistic regression for this problem.

Logistic Regression is used for :

1) Classification
2) Estimate / Assist risk
3) Establish relationships between explanatory variables (numerical / categorical) with categorical Y.

Logistic function : $f(x) = \dfrac{1}{1 + e^{-x}}$ $\quad\quad 0 \leq f(x) \leq 1$

$f(x) = \dfrac{e^{x}}{1 + e^{x}}$




$$\mathbb{P}(Y = 1 \mid X_1, \dots, X_p) = \frac{e^{\beta_0 + \sum_i \beta_i x_i}}{1 + e^{\beta_0 + \sum_i \beta_i x_i}} \quad , \quad Y \in \{0, 1\}$$

$$= \frac{1}{1 + e^{-(\beta_0 + \sum_i \beta_i x_i)}}$$ probability of

success

* Logit Transformation :

$$\text{logit}(P(X)) = \log\left(\frac{\mathbb{P}(Y=1 \mid X)}{1 - \mathbb{P}(Y=1 \mid X)}\right)$$

$\downarrow$

log odds of $\mathbb{P}(Y=1)$ vs $\mathbb{P}(Y=0)$

$$= \log\left(\frac{1}{1+e^{-(\beta_0+\sum \beta_i x_i)}} \quad \frac{1+e^{-(\beta_0+\sum \beta_i x_i}}{e^{-(\beta_0+\sum \beta_i x_i)}}\right)$$

$$= \log\left(e^{\beta_0+\sum \beta_i x_i}\right)$$

$$= \beta_0 + \sum_{i=1}^{n} \beta_i x_i$$

$$\text{odds} = \frac{\text{Prob of success}}{\text{Prob of failure}}$$

* Finding the parameters using MLE :

$$L = \exp\left\{\sum y_i \ln\left(\frac{p_i}{1-p_i}\right) + \sum n \ln(1-p_i) + \sum \ln\binom{n}{y_i}\right\}$$

$$\ln(L) = \prod_{i=1}^{n} \binom{n}{y_i} (1-p_i)^n \left(\frac{p_i}{1-p_i}\right)^{y_i},$$

$$= \prod_{i=1}^{n} \binom{n}{y_i} \left(1-(1+e^{\beta_0+\sum \beta_i x_i})\right) (\beta_0+\sum \beta_i x_i)^{y_i}$$

Numerical optimization tools will be used to compute of $\beta$.

Measure of Goodness : Imbalanced

## Classification Table

| Using Logistic Reg | Observed Values | | |
|---|---|---|---|
| | 0 | 1 | |
| 0 | 16 ✓ | 11 x | |
| 1 | 131 x | 417 ✓ | |
| | 147 | 428 | 575 |

Wrong Predictions : 131 + 11 = 142

0 : Having disease    1: Not having

Sensitivity = $\mathbb{P}$ ( false Predict disease ) = $\frac{11}{147}$

Classification Accuracy $= \frac{417 + 16}{575}$ if $> 0.8$ then it is a good model.

Assigning risk is done by logit!